A revised method of scoring

Vandaele, W.H.; Chowdhury, S.R.

Publication date:
1970

Document Version
Publisher's PDF, also known as Version of record

W. H. Vandaele and S. R. Chowdhury

A revised method of scoring

Research memorandum R 41

maximization

TILBURG INSTITUTE OF ECONOMICS
DEPARTMENT OF ECONOMETRICS

A Revised Method of Scoring

by

VANDAELE Walter H. and S. R.CHOWDHURY

1 Introduction

The "Method of Scoring" given by Fisher R.A.[†] is almost always suggested in statistical literatures to find out a relative maximum of the logarithm of the likelihood function when it can not be explicitly solved. Since it is an iterative procedure to find out a relative maximum, we would like to know about its convergence. Barnett V.D. [1] has pointed out, that the Method of Scoring (MS) may fail to converge, or even may converge to a relative minimum rather than to a relative maximum.

In this paper, the method is analysed from the principles of gradient method of maximization given by Crockett J.D. and H.B. Chernoff [2]. A simple modification is also suggested to ensure convergence to a relative maximum.

---

† FISHER, R.A. "Theory of statistical estimation", Proceeding of the Cambridge Philosophical Society. Vol. 22, 1925, pp. 700-725.

2 The Analysis of the Method of Scoring from Gradient Principle

Let $L_T(\theta)$ be the logarithm of the likelihood function of the parametervector $\theta = (\theta_1, \theta_2, \ldots\ldots, \theta_n)$ for a sample size T. Our problem is to find out a relative maximum of $L_T(\theta)$, for unrestricted θ, when direct methods fail to give an explicit solution. In this situation, starting with an initial approximation, an iterative method is usually applied to approximate the relative maximum reasonably well.

In order to examine the convergence in the Scoring Method, we should take a look at the steepest ascent or gradient principle.

As given by Crockett, J.B. and H. Chernoff [2], the iteration scheme for <u>Gradient or Steepest ascent method</u> is

$$\theta^{(i+1)} = \theta^{(i)} + h_i \; B^{-1} \; g^{(i)} \tag{2.1}$$

where:

- h_i is a positive scalar suitably chosen
- B is a positive definite matrix, being a weighting matrix
- $\theta^{(i)}$ is the value of the vector θ at the i'th iteration
- $g^{(i)}$ is the n-dimensional column vector of partial derivatives of $L_T(\theta)$ with respect to (w.r.t.) θ_i, evaluated at $\theta^{(i)}$

The gradient vector $B^{-1} \; g^{(i)}$ gives the direction of the steepest ascent at $\theta^{(i)}$ w.r.t. B; h_i is the length of the step taken in that direction. As we move from $\theta^{(i)}$ in the direction of $B^{-1} \; g^{(i)}$, $L_T(\theta)$ increases, i.e. a positive h_i can always be found such that

$$L_T(\theta^{(i+1)} = \theta^{(i)} + h_i \; B^{-1} \; g^{(i)}) > L_T(\theta^{(i)}) \tag{2.2}$$

The necessary condition that the iteration process will converge, and converge to a relative maximum is:

$$L_T(\theta^{(i+1)}) > L_T(\theta^{(i)}) \qquad \text{for each } i \qquad (2.3)$$

If the steps h_i are suitably chosen so as to satisfy (2.3), then the gradient method will always converge to a relative maximum. With this knowledge, let us now examine the Method of Scoring.

The iteration scheme in the <u>Method of Scoring</u> is given by

$$\theta^{(i+1)} = \theta^{(i)} + I^{(i)^{-1}} g^{(i)} \qquad (2.4)$$

with

$\theta^{(i)}$ and $g^{(i)}$ defined above

and

$$I^{(i)} = E\left[\frac{-\partial^2 L_T(\theta)}{\partial\theta_i \, \partial\theta_j}\right]_{\theta = \theta^{(i)}}$$: the information matrix at the i'th iteration.

Comparing (2.1) and (2.4), we find that the Scoring Method is in fact a Gradient Method, with $I^{(i)}$ replacing B, and the steps h_i being unity always. The matrix $I^{(i)}$ being a covariance matrix by formula, is always positive definite (see Kendall, M.G. and A.Stuart [6], pp. 35-74). By making steps equal to unity always, we are not sure whether condition (2.3) will be satisfied, and so we cannot definitely say whether the process will converge. It can even converge to a relative minimum, rather than a relative maximum. Thus, the usual Method of Scoring needs modifications to ensure convergence to a relative maximum.

3 A Revised Method of Scoring

A modified Method of Scoring will be given as

$$\theta^{(i+1)} = \theta^{(i)} + h_i \, I^{(i)^{-1}} g^{(i)} \tag{3.1}$$

(3.1) is different from (2.4) only in the steplength h_i, where h_i is defined in (2.1). The steplength h_i in (3.1) is chosen in such a way that (2.3) is always satisfied.

Selection of h_i

One way of choosing h_i and which is sometimes suggested, is to choose h_i such that $L_T(\theta^{(i+1)} = \theta^{(i)} + h_i I^{(i)^{-1}} g^{(i)})$ as a function of h_i, is a maximum. To find out a maximum, we have to solve for h_i

$$\frac{\partial L_T(\theta^{(i)} + h_i \, I^{(i)^{-1}} g^{(i)})}{\partial h_i} = 0 \tag{3.2}$$

If (3.2) could be explicitly solved i.e. if we know all the relative maxima and minima, then we have accomplished our purpose. We choose that h_i for which $L_T(\theta^{(i)} + h_i \, I^{(i)^{-1}} g^{(i)})$ is absolute maximum. In case we cannot solve (3.2) explicitly, we can try to find out the first relative extremum also by iteration. The first relative extremum will be a relative maximum, as the function $L_T(\theta^{(i)} + h_i \, I^{(i)^{-1}} g^{(i)})$ increases in the neighbourhood of $\theta^{(i)}$. As the first relative maximum will also satisfy (2.3), the process will converge to a relative maximum. To find out the first relative maximum of $L_T(\theta^{(i)} + h_i \, I^{(i)^{-1}} g^{(i)})$ w.r.t. h_i, we can start the itera-ation process with the initial value of h_i to be zero.

Note that, any relative maximum of $L_T(\theta^{(i)} + h_i \, I^{(i)^{-1}} g^{(i)})$ w.r.t. h_i may not do, for (2.3) may not be satisfied which is essential for convergence to a relative maximum.

Instead of trying to find h_i in the above way which requires much computations, we can adopt a simple procedure to find h_i such that (2.3) is satisfied. This practical procedure has been applied in the subsequent rapported examples.

A practical procedure

First take a unit step i.e. $h_i = 1$.

a) If $L_T(\theta^{(i)} + I^{(i)^{-1}} g^{(i)} > L_T(\theta^{(i)}$, then we go on doubling the steps until the first turning point is occurred.

$$L_T(\theta^{(i)} + I^{(i)^{-1}} g^{(i)}) > L_T(\theta^{(i)})$$

$$L_T(\theta^{(i)} + 2\, I^{(i)^{-1}} g^{(i)}) > L_T(\theta^{(i)} + I^{(i)^{-1}} g^{(i)})$$

$$L_T(\theta^{(i)} + n\, I^{(i)^{-1}} g^{(i)}) > L_T(\theta^{(i)} + \frac{n}{2} I^{(i)^{-1}} g^{(i)})$$

$$L_T(\theta^{(i)} + 2n\, I^{(i)^{-1}} g^{(i)}) < L_T(\theta^{(i)} + n\, I^{(i)^{-1}} g^{(i)})$$

In this case we take $\theta^{(i+1)} = \theta^{(i)} + n\, I^{(i)^{-1}} g^{(i)}$.

b) If $L_T(\theta^{(i)} + I^{(i)^{-1}} g^{(i)} < L_T(\theta^{(i)}$, we go on halving the steps until a turning point is reached.

$$L_T(\theta^{(i)} + I^{(i)^{-1}} g^{(i)}) < L_T(\theta^{(i)})$$

$$L_T(\theta^{(i)} + \tfrac{1}{2} I^{(i)^{-1}} g^{(i)}) < L_T(\theta^{(i)})$$

$$L_T(\theta^{(i)} + \frac{2}{n} I^{(i)^{-1}} g^{(i)}) < L_T(\theta^{(i)})$$

$$L_T(\theta^{(i)} + \frac{1}{n} I^{(i)^{-1}} g^{(i)}) > L_T(\theta^{(i)})$$

In this case we take $\theta^{(i+1)} = \theta^{(i)} + \frac{1}{n} I^{(i)^{-1}} g^{(i)}$.

In each of the cases a) and b), condition (2.3) is satisfied, and we are assured of the convergence to a relative maximum. Moreover, this procedure can reduce the number of iterations.

As example we will estimate the parameters in an Autocorrelated Model with the Revised Method of Scoring.

4 An Autocorrelated Model, an application of the Revised Method of Scoring

The Model is written in matrix notation as follows

$$\underline{y} = X\,\beta + \underline{u} \qquad (4.1)$$

where $\underline{y}$ is a column vector of T values taken by the dependent variable; X a matrix of order T x k of values taken by the k nonstochastic variables $x_1, \ldots\ldots, x_k$; β a column vector of k unknown parameters, and $\underline{u}$ a column vector of T nonobservable random variables, the disturbance.

The following assumptions about the vector of random variables and the X-matrix are made:

(i) The matrix X consists of nonstochastic elements and has rank $k \leq T$.

(ii) The random variables $\underline{u}_1, \ldots\ldots, \underline{u}_n$ are multinormally distributed.

(iii) The random variable is supposed to follow a first order autoregressive scheme:

$$\underline{u}_t = \rho\,\underline{u}_{t-1} + \underline{\varepsilon}_t$$

where $|\rho| < 1$ and $\underline{\varepsilon}_t$ has the following properties

$$- E(\underline{\varepsilon}_t, \varepsilon_{t+s}) = \left\{ \begin{array}{lll} \sigma^2 & \text{for} & s = 0 \\ 0 & \text{for} & s \neq 0 \end{array} \right\} \quad t = 1(1)T$$

$$- E(\underline{\varepsilon}_t) = 0$$

So

$$E\,\underline{u}\,\underline{u}' = V = \frac{\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\ \rho & & & & \rho^{T-2} \\ \vdots & & & & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots & 1 \end{bmatrix}$$

and

$$V^{-1} = \frac{1}{\sigma^2}\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & & 0 & 0 \\ \vdots & & & & & \vdots \\ & & & & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

It can easily be verified that $|V^{-1}| = \frac{1-\rho^2}{\sigma^{2T}}$ (4.2)

The Likelihood function of the sample is

$$L^* = \frac{|V^{-1}|^{\frac{1}{2}}}{(2\pi)^{T/2}} \exp\left[-\tfrac{1}{2}\ (y-X\hat{\beta})'\ V^{-1}\ (y-X\hat{\beta})\right] \tag{4.3}$$

Taking ln, defining $e \equiv y-X\hat{\beta}$ and inserting (4.2), (4.3) becomes:

$$\begin{aligned} L = \ln L^* &= -\frac{T}{2}\ln 2\pi + \tfrac{1}{2}\ln\left(\frac{1-\rho^2}{\sigma^{2T}}\right) - \tfrac{1}{2}e'V^{-1}e \\ &= -\frac{T}{2}\ln 2 + \tfrac{1}{2}\ln\left(\frac{1-\rho^2}{\sigma^{2T}}\right) \qquad (4.4) \\ &\quad - \frac{1}{2\sigma^2}\left[\sum_{t=1}^{T} e_t^2 + \rho^2\sum_{t=2}^{T-1} e_t^2 - 2\rho\sum_{t=1}^{T-1} e_t e_{t+1}\right] \end{aligned}$$

We have omitted T and θ in the notation $L_T(\theta)$. This will however not lead to any confusion.

In order to apply the Revised method of Scoring for estimation of σ, ρ and the β-vector, we have to build up the Scoringvector and Informationmatrix.

<u>Scoringvector</u>

The k + 2 components scoring vector is built up of

$$\begin{bmatrix} \frac{\partial L}{\partial \sigma} \\ \frac{\partial L}{\partial \rho} \\ \frac{\partial L}{\partial \beta_1} \\ \vdots \\ \frac{\partial L}{\partial \beta_k} \end{bmatrix}$$

where the components are the following algebraical expressions:

$$\frac{\partial L}{\partial \sigma} = -\frac{T}{\sigma} + \frac{1}{\sigma^3}\left[\sum_{t=1}^{T} e_t^2 + \rho^2 \sum_{t=2}^{T-1} e_t^2 - 2\rho \sum_{t=1}^{T-1} e_t e_{t+1}\right]$$

$$\frac{\partial L}{\partial \rho} = -\frac{\rho}{1-\rho^2} - \frac{1}{\sigma^2}\left[\rho \sum_{t=2}^{T-1} e_t^2 - \sum_{t=1}^{T-1} e_t e_{t+1}\right]$$

$$\frac{\partial L}{\partial \beta_i} = \frac{1}{\sigma^2}\left[\sum_{t=1}^{T} e_t x_{it} + \rho^2 \sum_{t=2}^{T-1} e_t x_{it} - \rho \sum_{t=1}^{T-1} (x_{it} e_{t+1} + e_t x_{i,t+1})\right]$$

$$i = 1,.., k.$$

<u>Informationmatrix.</u>

$$\frac{\partial^2 L}{\partial \sigma^2} = \frac{T}{\sigma^2} - \frac{3}{\sigma^4}\left[\sum_{t=1}^{T} e_t^2 + \rho^2 \sum_{t=2}^{T-1} e_t^2 - 2\rho \sum_{t=1}^{T-1} e_t e_{t+1}\right]$$

$$\frac{\partial^2 L}{\partial \sigma \partial \rho} = \frac{2}{\sigma^3}\left[\rho \sum_{t=2}^{T-1} e_t^2 - \sum_{t=1}^{T-1} e_t e_{t+1}\right]$$

$$\frac{\partial^2 L}{\partial \sigma \partial \beta_i} = -\frac{2}{\sigma^3}\left[\sum_{t=1}^{T} e_t x_{it} + \rho^2 \sum_{t=2}^{T-1} e_t x_{it} - \rho \sum_{t=1}^{T-1} (x_{it} e_{t+1} + e_t x_{i,t+1})\right]$$

$$\frac{\partial^2 L}{\partial \rho^2} = -\frac{(1+\rho^2)}{(1-\rho^2)^2} - \frac{1}{\sigma^2} \sum_{t=2}^{T-1} e_t^2$$

10.

$$\frac{\partial^2 L}{\partial\rho\,\partial\beta_i} = \frac{1}{\sigma^2}\left[2\,\rho \sum_{t=2}^{T-1} e_t\, x_{it} - \sum_{t=1}^{T-1} (x_{it}\, e_{t+1} + e_t\, x_{i,t+1})\right]$$

$$i = 1,\ldots,k$$

$$\frac{\partial^2 L}{\partial\beta_i\,\partial\beta_j} = -\frac{1}{\sigma^2}\left[\sum_{t=1}^{T} x_{it}\, x_{jt} + \rho^2 \sum_{t=1}^{T-1} x_{it}\, x_{jt} - \rho \sum_{t=1}^{T-1} (x_{it}\, x_{j,t+1} + x_{jt}\, x_{i,t+1})\right]$$

$$i,j = 1,\ldots,k$$

After taking expectation of the partial derivatives and multiplying with -1, we obtain the (k+2) x (k+2) Informationmatrix, the elements of which are

(1,1) $$-E\left(\frac{\partial^2 L}{\partial\sigma^2}\right) = -\frac{T}{\sigma^2} + \frac{3}{\sigma^4}\left[\frac{T\,\sigma^2}{1-\rho^2} + \rho^2\,(T-2)\,\frac{\sigma^2}{1-\rho^2} - 2\,\rho^2 (T-1)\,\frac{\sigma^2}{1-\rho^2}\right]$$

$$= \frac{2T}{\sigma^2}$$

(1,2) = (2,1) $$-E\left(\frac{\partial^2 L}{\partial\sigma\,\partial\rho}\right) = -\frac{2}{\sigma^3}\left[\rho\,(T-2)\,\frac{\sigma^2}{1-\rho^2} - \rho\,(T-1)\,\frac{\sigma^2}{1-\rho^2}\right]$$

$$= \frac{2\,\rho}{\sigma(1-\rho^2)}$$

(1,l) = (l,1)

l = 3,...,k+2 $$-E\left(\frac{\partial^2 L}{\partial\sigma\,\partial\beta_i}\right) = 0 \quad ; \quad i = 1,\ldots,k.$$

(2,2) $$- E\left(\frac{\partial^2 L}{\partial \rho^2}\right) = \frac{1}{1-\rho^2}\left[\frac{1+\rho^2}{1-\rho^2} + T - 2\right]$$

(2,1) = (1,2)

l = 3,...,k+2 $$- E\left(\frac{\partial^2 L}{\partial \rho \partial \beta_i}\right) = 0 \quad ; \quad i = 1,\ldots, k$$

The right-lower k x k symmetric matrix is:

$$- E\left(\frac{\partial^2 L}{\partial \beta_i \partial \beta_j}\right) = - \frac{1}{\sigma^2}\left[\sum_{t=1}^{T} x_{it}\, x_{jt} + \rho^2 \sum_{t=2}^{T-1} x_{it}\, x_{jt} - \rho \sum_{t=1}^{T-1} (x_{it}\, x_{j,t+1} + x_{jt}\, x_{i,t+1})\right] \quad i,j = 1,\ldots,k$$

So, the Information matrix looks like :

$$I = \begin{bmatrix} - E \frac{\partial^2 L}{\partial \sigma^2} & - E \frac{\partial^2 L}{\partial \sigma \partial \rho} & 0 \\ - E \frac{\partial^2 L}{\partial \rho \partial \sigma} & - E \frac{\partial^2 L}{\partial \rho^2} & 0 \\ 0 & 0 & - E \frac{\partial^2 L}{\partial \beta_i \partial \beta_j} \end{bmatrix}$$

Because of the particular structure of this Information matrix:

$$I = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \Longrightarrow I^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}$$

and as A is a (2 x 2) matrix, A^{-1} can be calculated analytically.

$$A = \begin{bmatrix} \dfrac{2T}{\sigma^2} & \dfrac{2\rho}{\sigma(1-\rho^2)} \\[2ex] \dfrac{2\rho}{\sigma(1-\rho^2)} & \dfrac{1}{(1-\rho^2)}\left[\dfrac{1+\rho^2}{1-\rho^2} + T - 2\right] \end{bmatrix}$$

$$|A| = \frac{2}{\sigma^2(1-\rho^2)}\left[\frac{T + \rho^2(T-2)}{1-\rho^2} + T^2 - 2T\right]$$

$$\text{Write } D \equiv \frac{T + \rho^2(T-2)}{1-\rho^2} + T^2 - 2T$$

Then

$$a^{11} = \left[\frac{1+\rho^2}{1-\rho^2} + T - 2\right] \sigma^2 / 2D$$

$$a^{12} = a^{21} = -\rho\sigma/D$$

$$a^{22} = T(1-\rho^2) / D$$

With the Information Matrix and Scoring vector defined, we have applied the Revised Method of Scoring on different examples where autocorrelation was present.† Two of this examples will be mentioned below. In the examples the usual Method of Scoring is also applied for comparison.

Here it can be stated that in examples where the Method of Scoring failed to converge, we obtained a solution by the RMS.

Remark 1.

Because in (4.4) the term $-\frac{T}{2}\log 2\pi$ is a constant part, we have only evaluated the $L_T(\theta)$ at each iteration by omitting that constant part. In the tables below, the value

† A discription of the computesprogram will be given in a subsequent paper.

of the $L_T(\theta)$ will invariably refer without that constant part.

Remark 2.

In all the examples we have started the iteration procedure with the least squares wstimates of σ, ρ and β-vector as initial values.

Example 1.

The data are generated from the from the following model

$$y_t = 3\,x_t + u_t$$

$$u_t = .5u_{t-1} + \varepsilon_t \qquad t = 1(1)15$$

where the ε's were drawn from a table of standardized random normal deviates. The x's are rescaled investment expenditures taken from a paper by Haavelmo, T, †. Alle figures are given in table (4.1).

Comparing tables (4.2) and (4.3) we may infer the following interesting points:

1° Both the methods have converged, the RMS in two, the MS in eighteen iterations. The computer time with the RMS is also much less than with the MS, which is expected.

2° The final values in the two methods are quite different. The final value of the last (without the constant part) in the usual MS is - 149.09468, being quite lower than the initial value obtained (- 11,19868). This suggests

---

† HAAVELMO, T. "Methods of Measuring the Marginal Propensity to Consume", Journal of the American Statistical Association. Vol. 42, 1947, pp. 105 - 122.

that with the Ms, we have most possibly obtained a relative minimum. With the RMS the final value of the lu L is higher than the initial one, as it should be, and has converged to a relative maximum. The estimates of the parameters by the RMS are ressonably near to the theoretical values whereas the usual MS is nowhere near the theoretical ones.

Table 4.1

Example 1: Haavelmo - model		
t	y_t	x_t
1	12.649	3.9
2	18.794	6.0
3	12.198	4.2
4	14.372	5.2
5	13.909	4.7
6	14.556	5.1
7	14.700	4.5
8	18.281	6.0
9	13.890	3.9
10	10.318	4.1
11	5.473	2.2
12	4.044	1.7
13	6.361	2.7
14	7.036	3.3
15	13.368	4.8

Table 4.2

Example 1:		Revised Method of Scoring			
Iteration number	Step-length	Value of $L_T(\theta)$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\beta}$
Initial value		- 11.19868	1.395	.345	2.928 (35.188)
1	1	- 16.09796	1.426	- .506	2.916
	1/2	- 12.43525	1.411	- .081	2.922
	1/3	- 11.50618	1.403	.132	2.925
	1/4	- 11.27293	1.399	.238	2.927
	1/5	- 11.21564	1.397	.291	2.928
	1/6	- 11.20207	1.396	.318	2.928
	1/7	- 11.19910	1.396	.331	2.928
	1/8	- 11.19858	1.396	.338	2.928
	1/9	- 11.19854	1.396	.341	2.928
2	1	- 16.16157	1.426	- .512	2.917
	1/2	- 12.46071	1.411	- .085	2.922
	⋮				
	0	- 11.19854	1.396	.341	2.928
FINAL VALUE		- 11.19854	1.396	.341	2.928 (35.185)

See Remark 1 and 2; Between brackets are the t-values of the regression coefficients.

Table 4.3

Example 1:		Method of Scoring		
Iteration number	Value of $L_T(\theta)$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\beta}$
Initial value	- 11.19868	1.395	.345	2.928 (35.188)
1	- 16.09796	1.426	- .506	2.916
2	- 26.62086	1.385	-1.139	2.931
3	- 15.75326	1.075	- .262	2.931
4	- 42.04496	1.055	-1.267	2.931
5	- 16.71441	.880	- .078	2.931
6	- 57.81633	.875	-1.266	2.930
7	- 21.43690	.720	- .003	2.931
8	- 81.38175	.720	-1.240	2.928
9	- 32.40521	.585	- .022	2.931
10	- 166.15147	.584	-1.182	2.929
11	- 59.60529	.465	- .156	2.931
12	- 150.18708	.461	- .960	2.930
13	- 149.39549	.462	- .958	2.931
14	- 149.17781	.462	- .958	2.931
15	- 149.11633	.462	- .958	2.931
16	- 149.09938	.462	- .958	2.931
17	- 149.09468	.462	- .958	2.931
18	- 149.09338	.462	- .958	2.931
FINAL VALUE	- 149.09468	.462	- .958	2.931 (103.333)

See Remark 1. and 2; Between brackets are the t-values of the regression coefficients.

Example 2.

The second example deals with the demand for textiles in the Netherlands from 1923 to 1939. The time series are given in table (4.4).

$$y_t = \beta_o + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \qquad t = 1(1)17$$

In this case y refers to the logarithm of consumption per head, x_1 to the logarithm of real income per head, and x_2 to the logarithm of the deflated price of the commodity. Hence β_o stands for the constant growth, β_1 for the income elasticity and β_2 for the price elasticity of textiles in the Netherlands in the period just mentioned.

The results of the Revised Method of Scoring and the simple Method of Scoring are presented in tables (4.5) and (4.6).

The example 2 gives the same type results as example 1. So we can draw the same conclusions as before.

An interesting feature is that in the spirits example used by Durbin, J. and G.S. Watson [†] the MS has even failed to converge, whereas the RMS has given consistent results.

† DURBIN, J. and G.S. WATSON. "Testing for Serial Correlation in Least Squares Regression II", Biometrika. Vol. 38, 1951, pp. 159 - 178.

Table 4.4

Example 2: Dutch Textile example.

t	y_t	x_{1t}	x_{2t}
1	1.99651	1.98543	2.00432
2	1.99564	1.99167	2.00043
3	2.00000	2.00000	2.00000
4	2.04766	2.02078	1.95713
5	2.08707	2.02078	1.93702
6	2.07041	2.03941	1.95279
7	2.08314	2.04454	1.95713
8	2.13354	2.05038	1.91803
9	2.18808	2.03862	1.84573
10	2.18639	2.02243	1.81558
11	2.20003	2.00732	1.78746
12	2.14799	1.97955	1.79588
13	2.13418	1.98408	1.80346
14	2.22531	1.98945	1.72099
15	2.18837	2.01030	1.77597
16	2.17319	2.00689	1.77452
17	2.21880	2.01620	1.78746

Table 4.5

Example 2:		Revised Method of Scoring					
Iteration-number	Step-length	Value of $L_T(\theta)$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
Initial-value		66.17695	.0135	- .101	1.373 (4.482)	1.144 (7.323)	- .829 (22.933)
1	1	66.19646	.0135	- .176	1.362	1.148	- .827
2	1	66.05178	.0135	- .262	1.366	1.147	- .828
	½	66.14571	.0135	- .219	1.364	1.147	- .828
	⋮						
	0	66.19646	.0135	- .176	1.362	1.148	- .827
FINAL VALUE		66.19646	.0135	- .176	1.362 (4.446)	1.148 (7.353)	- .827 (22.901)

See Remark 1. and 2; Between brackets are the t-values of the regression coefficients

Table 4.6

Example 2:	Method of Scoring					
Iteration number	Value of $L_T(\theta)$	$\hat{\sigma}$	$\hat{\rho}$	$\hat{\beta}_o$	$\hat{\beta}_1$	$\hat{\beta}_2$
Initial value	66.17695	.0136	- .101	1.373 (4.482)	1.144 (7.323)	- .829 (22.933)
1	66.19646	.0135	- .176	1.362	1.148	- .827
2	66.11846	.0135	- .256	1.354	1.151	- .827
3	65.98536	.0135	- .330	1.347	1.154	- .826
4	65.81350	.0135	- .396	1.342	1.156	- .826
5	65.62023	.0135	- .455	1.337	1.158	- .825
6	65.42039	.0135	- .505	1.334	1.159	- .825
7	65.22537	.0134	- .548	1.331	1.161	- .825
⋮						
45	63.951708	.0133	- .749	1.320	1.166	- .825
46	63.951597	.0133	- .749	1.320	1.166	- .825
47	63.951503	.0133	- .749	1.320	1.166	- .825
48	63.951400	.0133	- .749	1,320	1.166	- .825
49	63.951345	.0133	- .749	1.320	1.166	- .825
50	63.951290	.0133	- .749	1.320	1.166	- .825
51	63.951239	.0133	- .749	1.320	1.166	- .825
FINAL VALUE	63.951290	.0133	- .749	1.320 (4.405)	1.166 (7.634)	- .825 (23.340)

See Remark 1. and 2; Between brackets are the t-values of the regression coefficients

5. Conclusion.

The examples have shown that the Method of Scoring is not always reliable to pick out a relative maximum. It is also true that by adopting the simple practical procedure RMS as an improvement, we can avoid the pitfalls of the MS.

6. References.

{ 1 } BARNETT, V.D. "Evaluation of the maximum - likelihood estimator where the likelihood equation has multiple roots", Biometrika. Vol. 53, 1966, nrs 1/2, pp. 151 - 165.

{ 2 } CROCKETT, J.B. and H. CHERNOFF. "Gradient Methods of Maximization", Pacific Journal of Mathematics. Vol. 5, 1955, pp. 33-50.

{ 3 } FLETCHER, R. and M.J.D. POWELL. "A rapidly convergent descent method for minimization", Computer Journal. Vol. 6, 1963, pp. 163 - 168.

{ 4 } GREENSTADT, J. "On the relative efficiences of Gradient Methods", Mathematics of Computation. Vol. 5, July, 1967, nr. 99, pp. 360 - 367.

{ 5 } HARTLY, H.O. "The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares", Technometrics. Vol. 3, May, 1961, nr. 2, pp. 269 - 280.

{ 6 } KENDALL, M.G. and A. STUART. The Advanced Theory of Statistics, Vol. 2, Inference and Relationship. London, Charles Griffin, 1967, 2nd. ed., 690 pp.

E.I.T. 1970